Two-Stage Structure-Conditioned Vessel Segmentation and Keypoint Localization for DIEP Flap Planning

Chi Xu[1], Xuejian Li[2], Zhengguo Zhang[1], Xi Ouyang[2], Zhong Xue[2], and Dinggang Shen[1,2,3(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, 201210 Shanghai, China
[2] Department of Research and Development, United Imaging Intelligence, 200230 Shanghai, China
[3] Shanghai Clinical Research and Trial Center, 201210 Shanghai, China
Dinggang.Shen@gmail.com

Abstract. Automated CTA analysis for deep inferior epigastric perforator (DIEP) flap planning requires both vessel segmentation and variable-count perforator keypoint localization. Joint optimization of these tasks is challenging because their annotations differ substantially in density and supervision strength. We propose PerfSeg, a two-stage structure-conditioned framework, to first train and freeze a vessel segmentation network and then use the resulting vascular representations to guide keypoint localization. In particular, t he localization stage combines vessel and muscle–adipose interface features through Dual Spatial Cross Attention and suppresses anatomically implausible responses through Soft Gating. Experiments are conducted on 277 clinical CTA cases, which are divided into 221 training, 28 validation, and 28 held-out test cases. The selected vessel model achieves a clDice of 0.745 on the validation set. On the held-out test set, PerfSeg achieves a precision of 0.679 and a recall of 0.708 under one-to-one matching at the predefined 10-mm threshold. The mean localization error (MLE) among successfully matched pairs was 7.8 mm, while the symmetric Chamfer distance (CD) over the complete point sets is 17.6 mm. Compared with four alternative localization methods, PerfSeg achieves the highest precision and recall and the lowest CD. These results demonstrate the value of combining vascular and tissue-interface priors for anatomically constrained perforator localization.

Keywords: Breast Reconstruction · CTA · Vessel Segmentation · Keypoint Localization · Two-stage Learning.

1 Introduction

Deep inferior epigastric perforator (DIEP) flap breast reconstruction relies on accurate preoperative assessment of abdominal perforator vessels. Clinicians need

not only to evaluate vascular continuity but also to localize the emergence points where perforators traverse the rectus abdominis muscle and enter the subcutaneous layer [1, 7, 10, 14, 15]. Computed tomography angiography (CTA) provides high-resolution visualization of abdominal vascular anatomy and has been widely used for DIEP flap surgical planning [4, 5]. However, manual CTA interpretation is time-consuming and operator-dependent, motivating the development of automated methods for vascular structure analysis and perforator localization.

Automated DIEP CTA analysis involves two closely related tasks: vessel segmentation and perforator keypoint localization. Vessel segmentation aims to recover thin, highly branched vascular structures that are prone to discontinuities, whereas perforator localization aims to identify the positions where vessels emerge through the rectus abdominis muscle. Unlike conventional landmark detection tasks, which generally assume a fixed set and number of keypoints [6], the number of DIEP perforators varies across patients because of anatomical variability. Therefore, a model must accommodate a variable number of targets while maintaining high spatial localization accuracy. Moreover, perforator keypoints should lie on anatomically plausible vascular structures and near the muscle surface. Relying solely on local image features may therefore produce false responses away from the vessels.

Existing studies on automated DIEP analysis remain limited in dataset scale and annotation granularity and have primarily focused on vessel segmentation [3, 11]. Although these methods provide information about vascular trajectories, they generally do not explicitly localize the emergence points where perforators traverse the rectus abdominis muscle and enter the subcutaneous layer. Because the locations of these emergence points directly affect preoperative vessel selection and surgical route planning, segmented vascular maps alone provide insufficient actionable guidance for clinical practice.

To address these limitations, we propose PerfSeg, a structure-conditioned two-stage framework for vessel segmentation and variable-count perforator keypoint localization in DIEP CTA. We evaluate PerfSeg on 277 clinical CTA cases, split into 221 training, 28 validation, and 28 held-out test cases. The vessel segmentation backbone was selected exclusively on the validation set, where it achieves a clDice of 0.745, and is subsequently fixed for downstream keypoint localization. On the held-out test set, PerfSeg achieves a precision of 0.679 and a recall of 0.708 at a predefined 10-mm matching threshold, with a matched-pair localization error (MLE) of 7.8 mm and a set-level Chamfer distance (CD) of 17.6 mm.

PerfSeg comprises two stages: the first produces vessel segmentations and vascular structural features, while the second localizes perforator keypoints using these features together with muscle–adipose interface cues. To incorporate anatomical constraints, we introduce a Dual Spatial Cross Attention module that fuses vascular and interface features, and a Soft Gating module that suppresses responses in anatomically implausible regions. These components provide structural guidance during both feature learning and heatmap refinement, improving perforator detection and set-level localization.

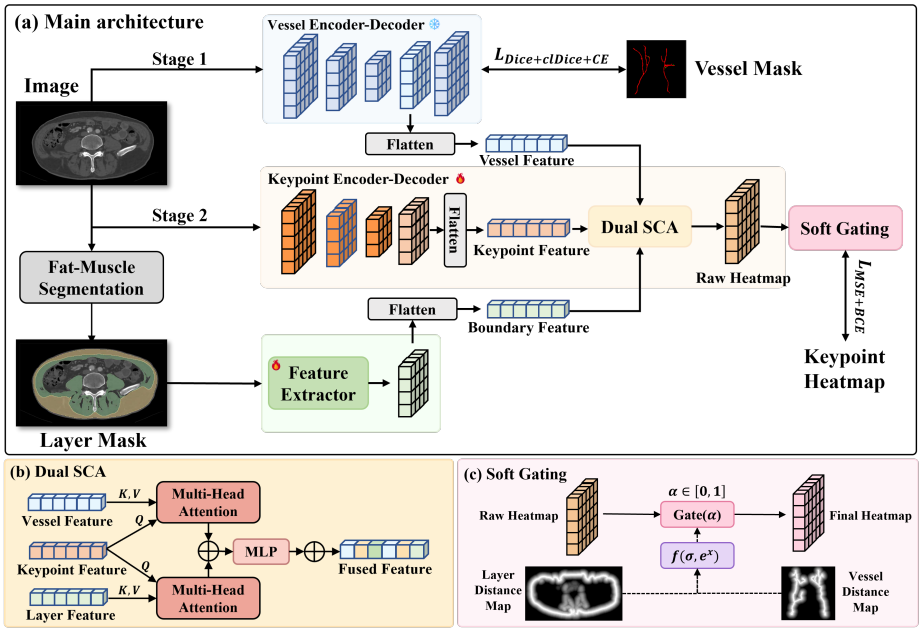

Fig. 1. Overview of the PerfSeg framework.

2 Method

The proposed framework adopts a concise two-stage design, as shown in Fig.1. In the first stage, an independent vessel segmentation model is trained to learn vascular structural representations from 3D CTA images; after training, this vessel segmentation model is frozen. In the second stage, a keypoint localization model is trained, where cross-attention is used during keypoint feature learning to jointly incorporate fat boundary information and vascular structural information, allowing both cues to guide the keypoint decision process. The model then predicts keypoint heatmaps, followed by a soft-gating module to filter keypoint responses far from the target location, thereby enhancing localization within anatomically plausible regions. The final output consists of the predicted keypoint locations together with the vessel segmentation result.

2.1 Dual Spatial Cross Attention (SCA)

The vessel feature $\mathbf{F}_v$ is extracted from the frozen first-stage vessel segmentation network, which is not updated by the keypoint-localization loss. To incorporate anatomical context, the Dual Spatial Cross Attention (Dual SCA) block cross-attends keypoint features $\mathbf{F}_k$ with vessel features $\mathbf{F}_v$ and muscle–adipose boundary features $\mathbf{F}_b$. These features are obtained from an intermediate decoder stage and flattened into token sequences of length $N = D'H'W'$, where (D', H', W') denotes the lower-resolution spatial dimensions.

The keypoint feature is used as the query, while the vessel and boundary features provide anatomical keys and values. Here, $\mathrm{MHA}(\cdot,\cdot,\cdot)$ denotes the multi-head attention operation, while $\mathbf{A}_v$ and $\mathbf{A}_b$ denote the vessel-aware and boundary-aware cross-attention outputs, respectively. The two cross-attention outputs are computed as

$$\mathbf{A}_v = \mathrm{MHA}(\mathbf{F}_k, \mathbf{F}_v, \mathbf{F}_v), \quad \mathbf{A}_b = \mathrm{MHA}(\mathbf{F}_k, \mathbf{F}_b, \mathbf{F}_b), \tag{1}$$

The vessel-aware and boundary-aware responses are then combined by summation and passed through a multi-layer perceptron (MLP) for feature modulation. A residual connection preserves the original keypoint representation:

$$\mathbf{F}_k^{\mathrm{out}} = \mathbf{F}_k + \mathrm{MLP}\left(\mathbf{A}_v + \mathbf{A}_b\right). \tag{2}$$

The resulting feature $\mathbf{F}_k^{\mathrm{out}}$ is therefore a mid-level keypoint representation conditioned on both vascular morphology and muscle–adipose boundary context. It is subsequently fed into the remaining decoder layers, which progressively upsample the feature to produce the final full-resolution keypoint heatmap.

2.2 Soft Gating

The Soft Gating Block suppresses anatomically implausible keypoint responses using vessel and tissue-interface priors. The vessel-derived prior map $D_v(\mathbf{x})$ is constructed from logits predicted by the frozen first-stage vessel segmentation network. Muscle and adipose masks are obtained using TotalSegmentator [16]. Both masks are dilated and their intersection is used to approximate the muscle–adipose interface. A Euclidean distance transform then produces the boundary distance map $D_b(\mathbf{x})$, with $D_b(\mathbf{x}) = 0$ on the interface.

Given $D_v(\mathbf{x})$ and $D_b(\mathbf{x})$, the spatial gate is defined as

$$G(\mathbf{x}) = \exp\left(-\frac{D_v(\mathbf{x})}{\sigma_v}\right) \cdot \exp\left(-\frac{D_b(\mathbf{x})}{\sigma_b}\right), \tag{3}$$

where σ_v and σ_b control the decay ranges of the vessel and boundary priors, respectively. The final heatmap is computed as

$$H_{\mathrm{final}} = H_{\mathrm{raw}} \odot \left(\alpha G + (1-\alpha)\right), \tag{4}$$

where α controls the gating strength. This residual gating preserves the original heatmap responses while down-weighting candidates distant from both the predicted vessel structure and the muscle–adipose interface.

2.3 Loss Functions

The network is optimized with two task-specific objectives for vessel segmentation and keypoint localization. For vessel segmentation, the loss combines cross-entropy (CE), Dice loss, and clDice loss:

$$\mathcal{L}_{seg} = \lambda_{CE}\mathcal{L}_{CE} + \lambda_{Dice}\mathcal{L}_{Dice} + \lambda_{clDice}\mathcal{L}_{clDice}. \tag{5}$$

Let $p_i \in [0,1]$ and $y_i \in \{0,1\}$ denote the predicted vessel probability and the ground-truth vessel label at voxel i, respectively. The CE loss is defined as

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right]. \tag{6}$$

The Dice loss is defined as

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i y_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \epsilon}, \tag{7}$$

where ϵ is a small constant for numerical stability.

For the clDice loss, let $S(p)$ and $S(y)$ denote the soft skeletons of the predicted vessel map and the ground-truth vessel mask, respectively. The topology-aware precision and sensitivity are computed as

$$T_{prec} = \frac{\sum_{i=1}^{N} S(p)_i y_i + \epsilon}{\sum_{i=1}^{N} S(p)_i + \epsilon}, \quad T_{sens} = \frac{\sum_{i=1}^{N} S(y)_i p_i + \epsilon}{\sum_{i=1}^{N} S(y)_i + \epsilon}. \tag{8}$$

The clDice loss is then defined as

$$\mathcal{L}_{clDice} = 1 - \frac{2T_{prec}T_{sens} + \epsilon}{T_{prec} + T_{sens} + \epsilon}. \tag{9}$$

For keypoint localization, the model predicts a heatmap q_i supervised by the target heatmap g_i. The keypoint loss is a weighted sum of mean squared error and binary cross-entropy:

$$\mathcal{L}_{kp} = \lambda_{\text{MSE}}\frac{1}{N_k}\sum_{i=1}^{N_k}(q_i - g_i)^2 - \lambda_{\text{BCE}}\frac{1}{N_k}\sum_{i=1}^{N_k}\left[g_i \log q_i + (1-g_i)\log(1-q_i)\right]. \tag{10}$$

3 Experiments

3.1 Dataset

This study uses a clinical DIEP CTA dataset comprising 277 cases. The dataset was divided at the patient level into 221 training cases, 28 validation cases, and 28 held-out test cases. The validation set was used for vessel-backbone selection, checkpoint selection, and other model-development decisions. After model selection, all configurations were fixed before evaluation on the test set. The test set was used only for final performance evaluation. Model selection and checkpoint selection were performed exclusively on the validation set. The held-out test set was evaluated only after all model configurations had been fixed.

ROI Cropping. We define a landmark-guided region of interest (ROI) in preprocessed image space from the bounding boxes of the L3 vertebra and hip bones segmented by TotalSegmentator. With the x-axis denoting the third volume dimension, the crop extends along x from the start of the hip-bone bounding box

to the end of the L3 bounding box plus a superior margin, along y from the minimum foreground coordinate to the start of the L3 bounding box plus a posterior margin, and along z over the hip-bone bounding box. Margins are specified in millimetres, converted to voxels using image spacing, and clipped to the image boundaries.

Normalization and Resampling. CTA intensities are clipped to $[-400, 1000]$ hounsfield units (HU) and linearly normalized to $[0, 1]$ as $x_{\text{norm}} = (x+400)/1400$. Images and masks are resampled to an isotropic spacing of 0.6 mm using cubic and nearest-neighbor interpolation, respectively.

Implementation Details. Unless otherwise noted, the proposed models are implemented in PyTorch and trained for 200 epochs using AdamW with a learning rate of 2×10^{-4} and a batch size of 8. Training uses automatic mixed precision with bfloat16 (BF16) on three NVIDIA H20 GPUs. During inference, MONAI sliding-window inference [2] generates full-volume predictions, and overlapping windows are aggregated by weighted averaging.

3.2 Evaluation Metrics

Vessel Segmentation. Centerline Dice (clDice) [13] is adopted as the primary evaluation metric to measure the topological agreement between predicted and ground-truth tubular structures. Volumetric Dice coefficient (Dice) is additionally reported to evaluate overall overlap between predicted and ground-truth segmentation masks. In addition, topological precision (T_{prec}) and topological sensitivity (T_{sens}) are reported, corresponding to centerline-based precision and recall, respectively, providing a fine-grained characterization of the completeness and accuracy of segmentation results.

Keypoint Localization. Predicted and reference keypoints are matched one-to-one in physical space using Hungarian matching based on Euclidean distance. At thresholds of $\tau = 5$ mm and $\tau = 10$ mm, representing strict and clinically acceptable criteria, matched pairs are counted as true positives, while unmatched predictions and reference points are counted as false positives and false negatives, respectively. We report precision, recall, and the mean localization error (MLE) over successfully matched pairs. As MLE is conditional on successful matching, it is interpreted together with precision and recall. We additionally report the threshold-independent symmetric Chamfer distance (CD) between the complete predicted and reference point sets.

4 Experimental Results

4.1 Vessel Segmentation

Table 1 compares vessel segmentation under supervision with or without auxiliary abdominal and lower-extremity vessel annotations, which provide additional vascular context beyond the bilateral DIEP vessels.

MedNeXt [12] achieves the best overall performance in both annotation settings, followed by nnU-Net [9], whereas Swin UNETR [8] performs worst.

Table 1. Validation-set comparison of vessel segmentation backbones and auxiliary supervision settings.

Auxiliary Setting	Method	Overlapping metrics			Topology metrics		
		Dice↑	Precision↑	Recall↑	clDice↑	Tprec↑	Tsens↑
w/o auxiliary vessels	nnU-Net [9]	0.655	0.594	0.730	0.689	0.690	0.668
	Swin UNETR	0.579	0.463	0.772	0.605	0.614	0.596
	MedNeXt [12]	0.684	0.668	0.701	0.703	0.798	0.640
w/ auxiliary vessels	nnU-Net	0.624	0.504	**0.822**	0.713	0.742	0.688
	MedNeXt	**0.699**	**0.670**	0.731	**0.745**	**0.808**	**0.692**

Table 2. Comparison with alternative perforator keypoint localization methods on the held-out test set. MLE is computed only over successfully matched pairs.

Method	$\tau = 5$ mm			$\tau = 10$ mm			CD↓
	Precision↑	Recall↑	MLE↓	Precision↑	Recall↑	MLE↓	
nnU-Net	0.348	0.455	3.6	0.473	0.651	5.7	24.3
MedNeXt	0.380	0.481	4.6	0.488	0.661	8.8	26.9
MedNeXt + Priors	0.443	0.506	3.1	0.508	0.663	**5.3**	20.1
Intersection Rule	0.471	0.656	**2.8**	0.572	0.698	6.4	21.2
PerfSeg	**0.608**	**0.664**	4.0	**0.679**	**0.708**	7.8	**17.6**

Auxiliary supervision improves MedNeXt from 0.684/0.703 to 0.699/0.745 in Dice/clDice. For nnU-Net, it improves clDice from 0.689 to 0.713 but reduces Dice from 0.655 to 0.624, indicating better vascular connectivity without improved voxel-level overlap. Based exclusively on validation performance, we select and freeze MedNeXt with auxiliary supervision for subsequent localization experiments.

4.2 Perforator Keypoint Localization and Module Analysis

Table 2 compares PerfSeg with four localization methods under the same data split and evaluation protocol: nnU-Net [9] and MedNeXt [12], adapted as unconditioned keypoint heatmap regressors; MedNeXt + Priors, which directly incorporates predicted vessel and tissue-interface maps without Dual SCA or Soft Gating; and Intersection Rule, a non-learning baseline that extracts centroids of 26-connected components from the intersection of predicted vessel and tissue-interface masks. All learning-based methods use identical preprocessing, target heatmaps, training schedules, and post-processing. PerfSeg achieves the highest precision and recall at both 5 mm (0.608/0.664) and 10 mm (0.679/0.708), as well as the lowest CD of 17.6 mm, compared with the lowest baseline CD of 20.1 mm. Although Intersection Rule and MedNeXt + Priors achieve lower conditional MLEs at 5 mm (2.8 mm) and 10 mm (5.3 mm), respectively, their lower detection rates and higher CD indicate weaker overall point-set localization.

Table 3. Ablation study of Dual SCA and Soft Gating for perforator keypoint localization on the held-out test set. MLE is computed only over successfully matched pairs.

Modules		Threshold-based metrics						CD↓
Dual SCA	Soft Gating	$\tau = 5$ mm			$\tau = 10$ mm			
		Precision↑	Recall↑	MLE↓	Precision↑	Recall↑	MLE↓	
		0.380	0.481	4.6	0.488	0.661	8.8	26.9
✓		0.476	0.529	4.8	0.560	0.672	9.0	25.3
	✓	0.558	0.514	**3.5**	0.611	0.672	**7.2**	19.8
✓	✓	**0.608**	**0.664**	4.0	**0.679**	**0.708**	7.8	**17.6**

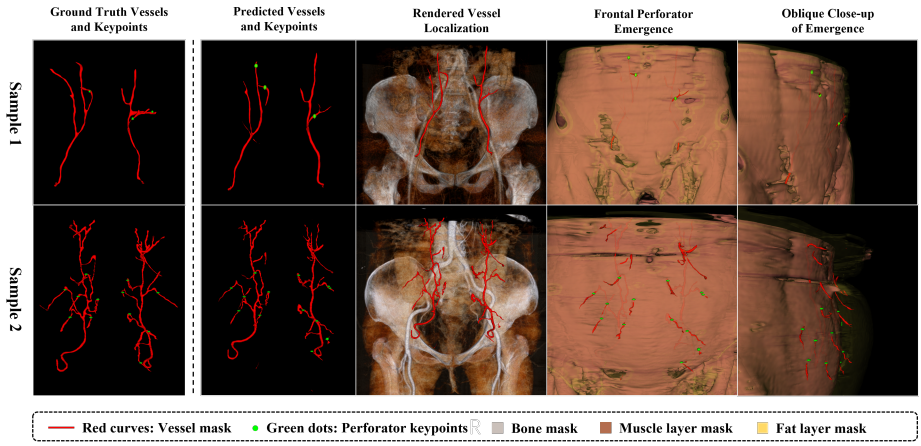

Fig. 2. Qualitative visualization of vessel segmentation and perforator localization. For each case, the left panel shows the ground truth, and the four right panels show the predicted mask, rendered vessel localization, frontal emergence view, and oblique close-up of the perforator emergence point.

The ablation results suggest complementary roles for the two modules. Dual SCA improves 5-mm precision/recall from 0.380/0.481 to 0.476/0.529, with a slight increase in MLE and a similar trend at 10 mm. Soft Gating reduces MLE to 3.5 mm and 7.2 mm at 5 mm and 10 mm, respectively, and reduces CD from 26.9 mm to 19.8 mm. Combining both modules achieves the highest precision and recall at both thresholds and the lowest CD. Overall, Dual SCA mainly improves detection coverage, whereas Soft Gating improves anatomical filtering and set-level localization.

Figure 2 presents representative localization results. Red curves denotes segmented DIEP vessels and green markers denotes predicted perforator keypoints. The predicted keypoints lie on vascular branches and remain spatially consistent with the abdominal wall surface, supporting the expected anatomical relationship between perforator trajectories and their emergence locations.

5 Conclusion

PerfSeg couples vessel segmentation with variable-count perforator keypoint localization through a two-stage structure-conditioned framework. The frozen vessel representation, muscle–adipose interface features, and soft anatomical gating jointly constrain keypoint predictions toward plausible perforator emergence regions. On the held-out test set, PerfSeg achieves a precision of 0.679 and a recall of 0.708 at the predefined 10-mm matching threshold, together with an MLE of 7.8 mm among successfully matched pairs and a CD of 17.6 mm over the complete point sets. These results demonstrate a favorable balance between perforator detection and set-level localization. The findings provide a basis for automated DIEP CTA analysis, while independent external validation remains necessary before clinical deployment.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Allen, R.J.: Diep versus tram for breast reconstruction. Plastic and reconstructive surgery **111**(7), 2478 (2003)
2. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Ceranka, J., Lamtenzan, D., Boonen, P., Kapila, A., Brussaard, C., Hamdi, M., Vandemeulebroucke, J.: Computer-assisted surgical planning for diep flap breast reconstruction surgery. In: Deep Breast Workshop on AI and Imaging for Diagnostic and Treatment Challenges in Breast Care. pp. 123–133. Springer (2025)
4. Daly, L.T., Doval, A.F., Lin, S.J., Tobias, A., Lee, B.T., Dowlatshahi, A.S.: Role of cta in women with abdominal scars undergoing diep breast reconstruction: review of 1,187 flaps. Journal of Reconstructive Microsurgery **36**(04), 294–300 (2020)
5. Figus, A., Canu, V., Iwuagwu, F.C., Ramakrishnan, V.: Diep flap with implant: a further option in optimising breast reconstruction. Journal of plastic, reconstructive & aesthetic surgery **62**(9), 1118–1126 (2009)
6. Geng, Z., Sun, K., Xiao, B., Zhang, Z., Wang, J.: Bottom-up human pose estimation via disentangled keypoint regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14676–14686 (2021)
7. Guerra, A.B., Metzinger, S.E., Bidros, R.S., Rizzuto, R.P., Gill, P.S., Nguyen, A.H., Dupin, C.L., Allen, R.J.: Bilateral breast reconstruction with the deep inferior epigastric perforator (diep) flap: an experience with 280 flaps. Annals of Plastic Surgery **52**(3), 246–252 (2004)
8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Lecture Notes in Computer Science, vol. 12962, pp. 272–284. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_26
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)

10. Keys, K.A., Louie, O., Said, H.K., Neligan, P.C., Mathes, D.W.: Clinical utility of ct angiography in diep breast reconstruction. Journal of Plastic, Reconstructive & Aesthetic Surgery **66**(3), e61–e65 (2013)
11. Moes, I.: Segmentation of deep inferior epigastric artery perforators for breast reconstruction using deep learning in computed tomography angiography: A quantitative and qualitative evaluation (2023)
12. Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H.: Mednext: transformer-driven scaling of convnets for medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 405–415. Springer (2023)
13. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16560–16569 (2021)
14. Taghizadeh, R., Moustaki, M., Harris, S., Roblin, P., Farhadi, J.: Does post-mastectomy radiotherapy affect the outcome and prevalence of complications in immediate diep breast reconstruction? a prospective cohort study. Journal of Plastic, Reconstructive & Aesthetic Surgery **68**(10), 1379–1385 (2015)
15. Teunis, T., Van Voss, M.H., Kon, M., van Maurik, J.M.: Ct-angiography prior to diep flap breast reconstruction: a systematic review and meta-analysis. Microsurgery **33**(6), 496–502 (2013)
16. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., et al.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)